Structure and Flow of Data Through the LLNL/ERD Environmental Data Management System

by Suzie Chamberlain

Introduction

This section will describe the structure and flow of data in the data management system used by the Environmental Restoration Division (ERD) of the Environmental Protection Department (EPD) of Lawrence Livermore National Laboratory (LLNL) to store and archive data from the Dynamic Stripping (DUS) Project.

The structure is based on two relational databases that are closely correlated. The first database handles sample tracking. The second database handles data such as sample location, media, analytical results, and some geological information. See Figure 1. These two databases are maintained on a VAX 6310 with VMS using INGRES relational database software. The flow of data, both hard copy and electronic, follows a model which tracks information from sampling plan through storage to archiving. The steps in the process include chain-of-custody (CoC) tracking of the sample, analytical result receipt, the application of quality control procedures, and the electronic use of data in decision support tools, such as risk assessment and compliance monitoring.

Structure and Flow

A sample plan is developed to establish the frequency, method and location of samples to be taken (see Figure 2). Field log books and CoC forms confirm the collection of samples according to the plan. A document control number is assigned to the samples based on the field log book used. A carefully controlled system of field log book labeling permits electronic tracking of an environmental sample from field collection through analytical result receipt as well as tracing back to the log book for any given analyte, should details of sampling conditions be needed. Samples are sent on to analytical laboratories where they are given unique log numbers. The important fields in each SPACT record are document control number, analytical laboratory, analytical lab log number, sampling location identification, sampling date, and the analysis requested. Several dates tracked are: receipt of CoC form and analytical results, and date of entry. SPACT also includes invoice information. Receipt of official printed analytical results and invoices cause database records to be updated based on the document control number and location. A data record is marked complete only when all analytical results have been received. Thus, completion of a record confirms that all requested analyses have been performed and reported.

Analytical results are stored in a separate, but correlated, relational database based on sampling location, log number, and date. This database is called MONITOR(Figure 3) and is related to SPACT by identical fields: document control number, sampling location, sampling date, analytical laboratory and requested analysis. Additional information collected for each sample and analyte includes requester, project, sample media, sample type, units, error, detection limit, dilution factor, and dates of extraction, analysis, and entry, together with comments and

special notes. Sources of data in this database include geologic borehole logs, surveyor reports, field measurements, laboratory measurements, calculated or reduced data, and test conclusions. Types of data to be stored have included descriptive sample location information, such as coordinates, elevations, lithology, and screened intervals of monitoring installations, as well as measurements and analytical information, including physical and chemical parameters, media identification, and ground water elevation measurements.

Data verification and validation are achieved through a combination of methods. Hand entered data are run through a series of computerized verifications that check for duplication, empty fields, and reported results not consistent with reported detection limits. Data are also thoroughly checked by a second person before being formally added to the database. Electronically delivered laboratory data are groomed to fill in empty fields and insure consistency in fields such as sample location, project, media, and type. Computerized verifications are also run on electronic data and a second person checks sample descriptor fields before formal addition to the database. Random audits are done to verify electronically delivered results against official printed results. Analytical results added to the database are validated through review by qualified chemists.

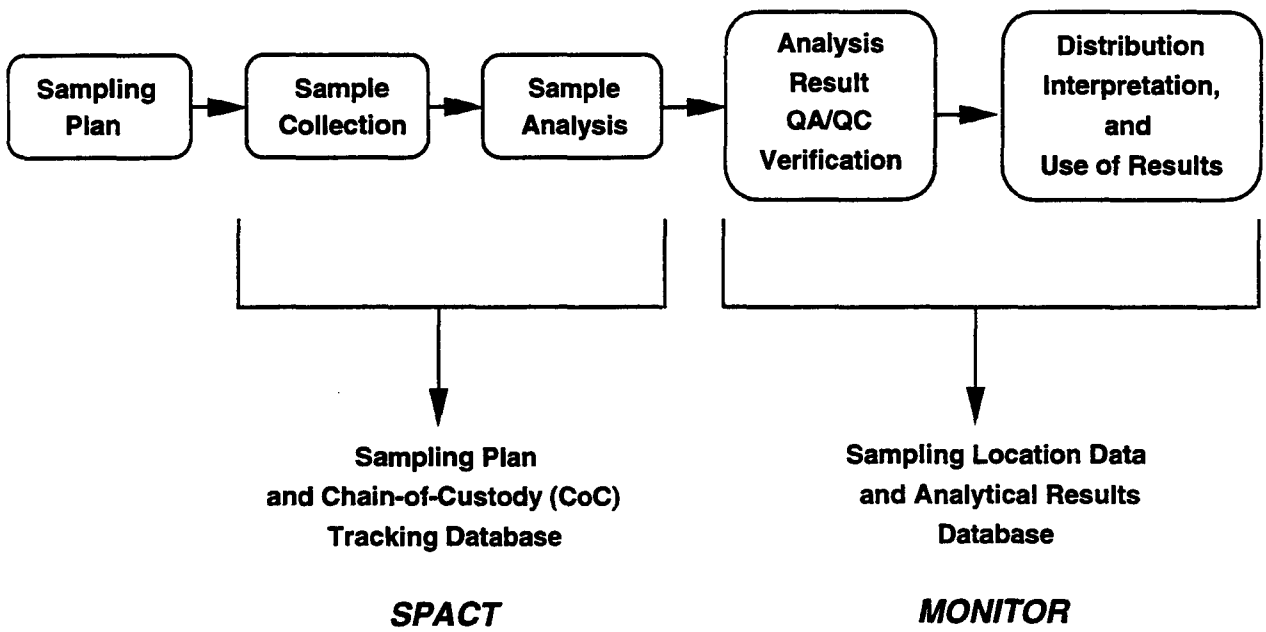
The MONITOR database also contains fields dedicated to quality control. Such fields include flags indicating analytical result qualification and data quality level. The result qualifier flags are absent from a routine report, but may be included to show non-detection, dilution greater than 1, rejection, or any of several other types. Data quality levels can range from EPA approved methods performed by a certified laboratory to quick, approximate field analyses.

Printed copies of the data received are filed by location and chronological order in a data reference library. The originals are retained for data submittal to regulators when required.

The operations data are stored in spreadsheets on Macintosh computers at the Gas Pad. Complete backups on this data occur on a weekly basis with incremental backups occurring daily. These weekly backups are currently stored in T4383. Plans are being developed for fire resistant storage. To allow for long term access to the backed up data, each month a tape is marked to save until data generation ceases. At the time data generation ceases, three complete archival backups will be performed and be stored in separate locations. This procedure will provide the highest level of availability of the data in the future.

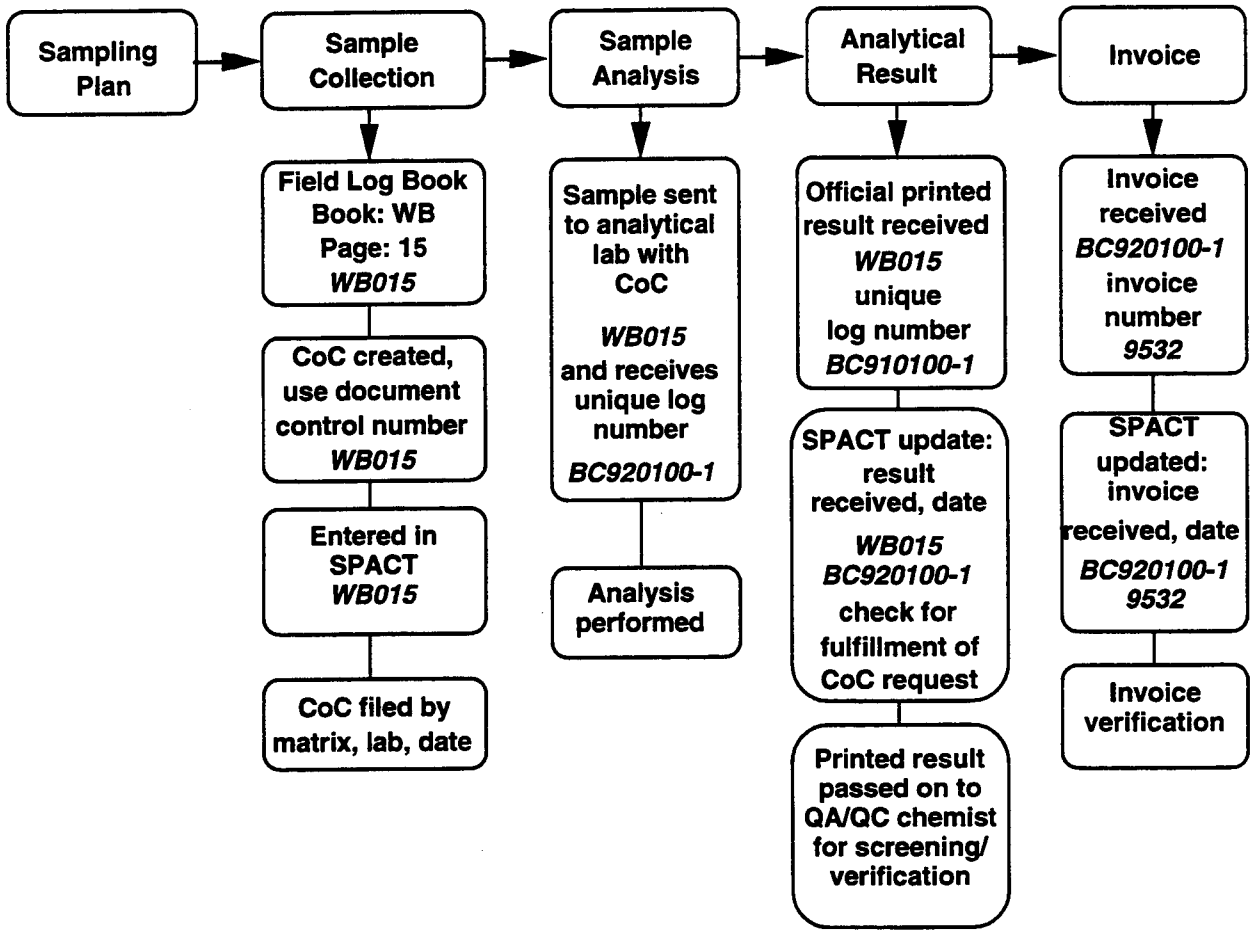
Conclusions

There are many advantages of this integrated centralized data management system. The use of such a system promotes and provides a consistent data set of known quality, which is available to all. Single entry for multiple use allows quality assurance and quality control to be performed equally for all data. Once an error is discovered and corrected, the data will be correct for all future uses. Interesting correlations and trends in the data may become obvious from linked extractions in the future.



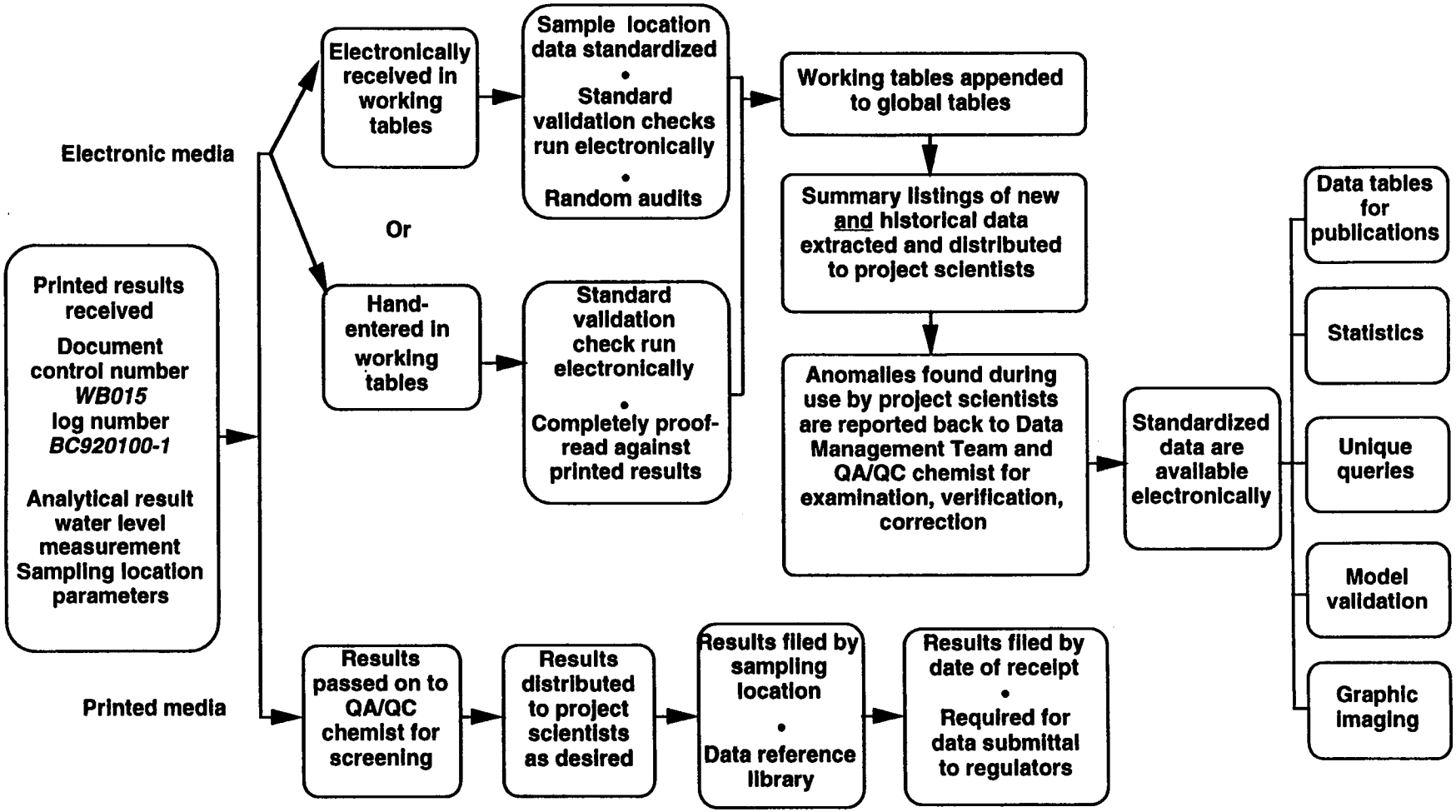
ERD-LSR-93-0027

Figure 1. Environmental data flow.



ERD-LSR-93-0028

Figure 2. Environmental data flow - SPACT Database.



ERD-LSR-93-0029

Figure 3. Environmental data flow - MONITOR database.